OAEI 2016 Results of AML

Daniel Faria[1], Catia Pesquita[2], Booma S. Balasubramani[3], Catarina Martins[2], João Cardoso[4], Hugo Curado[2], Francisco M. Couto[2], and Isabel F. Cruz[3]

[1] Instituto Gulbenkian de Ciência, Portugal
[2] LaSIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal
[3] ADVIS Lab, Department of Computer Science, University of Illinois at Chicago, USA
[4] INESC-ID, Instituto Superior Técnico, Universidade de Lisboa

Abstract. AgreementMakerLight (AML) is an automated ontology matching system based primarily on element-level matching and on the use of external resources as background knowledge. This paper describes its configuration for the OAEI 2016 competition and discusses its results.
For this OAEI edition, we tackled instance matching for the first time, thus expanding the coverage of AML to all types of ontology matching tasks. We also explored OBO logical definitions to match ontologies for the first time in the OAEI.
AML was the top performing system in five tracks (including the Instance and instance-based Process Model tracks) and one of the top performing systems in three others (including the novel Disease and Phenotype track, in which it was one of three prize recipients).

1 Presentation of the System

1.1 State, Purpose, General Statement

AgreementMakerLight (AML) is an automated ontology matching system derived from AgreementMaker [3, 4] and designed to tackle large-scale matching problems [6]. It is based primarily on lexical matching techniques, with an emphasis on the use of external resources as background knowledge.
This year, our development of AML was focused primarily on tackling instance matching, an aspect of ontology matching that was missing from its portfolio. However, we also made several developments with regard to class matching, namely with the use of OBO logical definitions.
For this OAEI edition, we also decided to adopt the solution of using configuration files for each track in order to specify the parameters of the matching task (such as whether to match classes, properties, and/or instances) rather than submit a preconfigured system. With this, we aim at providing a more transparent approach to our participation in the OAEI.

1.2 Specific Techniques Used

For the sake of brevity, this section describes only the features of AML that are new for this edition of the OAEI. For a complete description of AML's matching strategy,

please refer to last year's OAEI results paper [5].

1.2.1 Ontology Data

To store data about ontology individuals, we expanded AML's *Lexicon* and *RelationshipMap* data structures [6] and created the new *ValueMap*. The current organization of these data structures is the following:

- The *Lexicon* of each *Ontology* stores local names (if not alpha-numeric codes), labels and other lexical annotations of classes, individuals, and properties, after normalizing them.
- The *ValueMap* of each *Ontology* stores all other annotations of individuals and their data property values.
- The global *RelationshipMap* stores relations between classes, between individuals, between properties, classes instanced by individuals, and property domains and ranges.

1.2.2 Instance Matching

For instance matching, AML's core strategy consists of three matching algorithms:

- The *HybridStringMatcher* which matches two entities by computing the maximum of the string similarity, word similarity, and WordNet similarity between their *Lexicon* entries. It is the algorithm AML already used to match properties.
- The *ValueStringMatcher* which matches two individuals by computing the maximum string similarity between their *ValueMap* entries, penalizing matches where the annotation or data property is not the same.
- The *Value2LexiconMatcher* which employs the same combination of similarity metrics as the *HybridStringMatcher*, but compares *Lexicon* entries of one entity with *ValueMap* entries of the other and vice versa.

AML's similarity score is the maximum of these three algorithms, but it uses a linear combination of the three to break similarity ties when performing alignment selection. AML deviates from this core strategy in three circumstances:

- When the matching problem requires translation, in which case it employs the same matching strategy used for classes and properties when translation is involved.
- When the ontologies have a high individual connectivity (indicating that there is a network or pipeline of individuals), in which case it employs the *ProcessMatcher* algorithm that was developed for matching business process models [2]. It combines string similarity with structural similarity.
- When the fraction of individuals with exactly matching values in the ontologies is high (meaning that matches based on values have low significance), in which case it employs only the *HybridStringMatcher*.

1.2.3 Exploring OBO Logical Definitions

OBO [12] logical definitions (or cross-products) provide definitions of ontology classes by establishing intersections between other classes, typically from different ontologies.

For example, the logical definition of the class Human Phenotype Ontology (HP) [10] class HP:0005815 (“supernumerary ribs”) corresponds to an intersection of the Phenotypic Quality Ontology class PATO:0002002 (“has extra parts of type”) and the Foundational Model of Anatomy (FMA) class [11] FMA:7574 (“rib”) via an ‘inheres_in’ relation. We had previously developed a variant of AML for computing this type of compound mapping [8].
For this year’s OAEI, we explored the use of these logical definitions to match ontologies that contain them. Continuing the previous example, the Mammalian Phenotype Ontology (MP) [13] contains the class MP:0000480 (“increased rib number”) which to an English-speaking human should be obvious that it corresponds to the HP class above. However, to lexical ontology matching algorithms this correspondence is very hard to detect. Logical definitions can help us find this mapping, as MP defines that the class above corresponds to an intersection of the same class PATO:0002002 and the UBERON class [7] UBERON:0002228 (“rib”). As UBERON has cross-references to FMA, we can automatically establish a correspondence between UBERON:0002228 and FMA:7574, and thus find the mapping HP:0005815 <=> MP:0000480.
Because the versions of HP and MP used in the OAEI didn’t include the logical definitions in the ontology files (as the versions available at the OBO portal do), we used an external file containing these definitions as background knowledge.

1.2.4 Thesaurus Matching

For this year’s OAEI we also employed a matching algorithm based on a thesaurus that is automatically derived from the ontologies by comparing labels and synonyms for the same classes, as we have described in a previous study [9]. We hadn’t used this strategy in previous OAEI editions because our original implementation was too broad and consequently both too imprecise and too inefficient computationally. We addressed these problems by making a more restrictive implementation.
Currently, the algorithm infers synonyms to populate the thesaurus only when two *Lexicon* entries for a class have the same number of words and all their words are equal except for one, in which case the words in which they differ are inferred to be synonymous. Additionally, the new *Lexicon* entries generated for classes using the thesaurus are now only used to check for literal full-name matches, whereas previously they were also used with string similarity algorithms.

1.3 Adaptations made for the evaluation

The adaptations made for the evaluation were: the preprocessing of cross-references from Uberon and DOID for use in the Anatomy and Large Biomedical Ontologies tracks, due to namespace differences; the use of an external logical definitions file, due to the absence of these in the versions of the ontologies used in the Disease and Phenotype track; and the precomputing of translations, due to Microsoft® Translator’s query limit.

1.4 Link to the system and parameters file

AML is an open source ontology matching system and is available through GitHub (https://github.com/AgreementMakerLight) as an Eclipse project, as a stand-alone Jar application, and as a package for running through the SEALS client.

2 Results

2.1 Anatomy

Thanks to the use of the new thesaurus matching algorithm, AML improved both its recall and recall+ to the highest ever results in this track (93.6% and 83.2% respectively). However, it had a 0.6% drop in precision and a 0.1% drop in F-measure in comparison with last year. It remains the best performing system in this track.

2.2 Benchmark

As in previous years, AML obtained a very high precision in this track (this year the highest, at 100%) but a low recall (0.24%) and consequently a low F-measure as well (38%). We maintain AML focused on matching real-world ontologies, and have not prioritized the Benchmark track.

2.3 Conference

AML's performance in the Conference track was exactly the same as last year, as the new developments do not affect its performance in this track. It remains the best performing system overall in this track, with the highest F-measure on the full reference alignment 1 (74%), on the full reference alignment 2 (70%, tied with CroMatch), and on both evaluation modalities with the uncertain reference alignment (Discrete: 78%; Continuous: 77%).
Concerning the logical reasoning evaluation, AML again had no consistency principle violations, but did have conservativity principle violations as this is an aspect AML deliberately doesn't take into account given that many of these violations are false positives.

2.4 Disease and Phenotype

AML was considered one of the three top systems in the Disease and Phenotype track. In the HP-MP task, it obtained F-measures of 86% and 89.7% according to the 2-vote and 3-vote silver standards, respectively, and produced 122 unique mappings with 86.7% precision. In the DOID-ORDO task, it obtained F-measures of 90.8% and 87.5% according to the 2-vote and 3-vote silver standards, respectively, and produced 308 unique mappings, with an estimated precision of 86.7%. AML's performance in capturing the manually created mappings was poorer (75.9% and 0% recall, for HP-Mp and DOID-ORDO respectively), since the majority of these mappings are subsumption ones and AML focuses on equivalence matching.

2.5 Instance Matching

In the Sabine sub-track, AML obtained the second highest F-measure in the Sabine Linguistic task, with 91.8%, and the highest F-measure in the Sabine Linking task, with 88.9%.
In the Synthetic sub-track, AML obtained the highest F-measure in the UOBM mainbox task, with 51.2%, and the second highest F-measure in the SPIMBENCH mainbox task, with 81.6%. Interestingly, it ranked lower on the corresponding sandbox versions (second in UOBM with 66.5%, and third in SPIMBENCH with 82%) and was the system that lost the least performance between the sandbox and the mainbox tasks. Additionally, it is important to mention that AML does not process or attempt to match individuals without class assignment, and that there were a number of these in both the UOBM and SPIMBENCH ontologies which were supposed to be matched, which resulted in lower scores for AML.
In the Doremus sub-track, AML obtained the highest F-measure in all three tasks, with 91.8% in the 9 heterogeneities task, 84.8% in the larger 4 heterogeneities task, and 88.60% in the false-positive track task.
Overall, AML obtained the top F-measure in five of the seven Instance Matching tasks, and second in the other two, making it overall the most successful instance matching system in the OAEI 2016.

2.6 Interactive Matching

AML had a worse performance than last year in this track, due to changes to its user interface to enable alignment revision, which affected the internal functioning of the interactive matching algorithm. We were unable to completely solve this issue in time for the evaluation. Nevertheless, in the Anatomy dataset, AML still had the highest F-measure (95.8% with 0% errors), the lowest number of oracle requests, and the lowest impact of errors, with a drop in performance under 3% between 0 and 30% errors. In the Conference dataset, it was surpassed by Alin in F-measure and by LogMap with regard to the lowest number of requests and lowest impact of errors.

2.7 Large Biomedical Ontologies

Like in the Anatomy track, the introduction of the thesaurus matching algorithm led to an improved recall from AML on the Large Biomedical Ontologies track, and as a result AML had a higher F-measure overall in all tasks than in previous years. Despite this, it was surpassed in F-measure on the FMA-NCI small and FMA-SNOMED small tasks, obtaining only the second-highest F-measure (ignoring the XMAP results, since this system uses the UMLS metathesaurus as background knowledge, which is the basis of the reference alignments). Nevertheless, it remains the best performing system able to complete all the tasks of this track, and the one that produces the most coherent alignments.

2.8 Multifarm

AML obtained the top F-measure when matching the same ontologies, and the third best when matching the same ontologies, due to lowered recall. Despite not being a systems specifically targeting cross-lingual matching, by using a translation module AML is able to achieve a good ranking in performance in this track.

2.9 Process Model

AML obtained the top F-measure result in this track, with 70.2%, surpassing not only all other ontology matching systems, but also all process model matching systems from last year's process model matching competition [1].

3 General comments

3.1 Comments on the results

AML remained among the top performing systems in nearly all preexisting tracks, while also obtaining top results in the new tracks: Disease and Phenotype, in which it was one of the prize winners; Process Model, in which it surpassed the results of (non-ontology) process model matchers; and Instance Matching with all new datasets. It was also consistently among the fastest systems and among those that produced the most coherent alignments. These results reflect our continued effort to extend AML to cover all types of ontology matching tasks while ensuring that it remains both effective and efficient.

3.2 Comments on the OAEI test cases

We welcomed the efforts to expand the scope of OAEI with new tracks and improve existing ones. We take this opportunity to highlight some issues we encountered during this year's competition, and suggest some possible improvements for future editions.
This year there were several issues with the test cases from the Instance Matching track: there were encoding problems associated with the Sabine datasets; there were instances without class assignments in the Synthetic and Doremus datasets, and in the case of the former, some of these instances were supposed to be matched; and the target ontology in the SPIMBENCH mainbox dataset was inconsistent. These are all issues that can be found in real-world datasets, and both the developers and users of ontology matching systems should be aware of them, but we believe that asking systems to handle such specific issues involves a high level of manual work and tuning of the systems, making their comparison less straightforward and transparent.
We also find that the evaluation in the Disease and Phenotype track still has room for improvement. Generating silver standards from the alignments produced by the participating systems via voting is a reasonable starting point for producing a reference alignment, but an insightful evaluation would then need that the silver consensus standards be manually validated, as well as the unique mappings produced by each system. Since only the latter manual evaluation was done, and for only up to 30 mappings, this

distorts the results as the evaluation will include wrong mappings (that multiple systems get wrong) and miss correct mappings (that only one system finds). Additionally, we propose that in next years the versions of the HP and MP ontologies used in this track include logical definitions, so other systems can also explore them.

4 Conclusion

In 2016, AML was the top performing system in five tracks (Anatomy, Conference, Instance, Multifarm, and Process Model) and one of the top performing systems in three others (Disease and Phenotype, Interactive, and Large Biomedical Ontologies). It fully met our goals and expectations for this year's competition, and rewarded our investment in instance matching (with top results in both Instance and Process Model) and our use of logical definitions (with a prize in the Disease and Phenotype track).
Nevertheless we remark with enthusiasm on the improvement of other matching systems in tracks such as Anatomy, Conference, and Large Biomedical Ontologies. While in previous years we could be led to the conclusion that ontology matching was stagnating, and that surpassing the results of the top systems would be a tall order, the results of this year's OAEI show that that is not the case.

Acknowledgments

The authors are thankful to Daniela Oliveira (Insight Centre for Data Analytics, NIU Galway, Ireland) for her support in the alignment of phenotype ontologies, and to André Oliveira, Filipa Marques and Tânia Maldonado, for their contribution to analyzing the test cases and results. FMC, CM and CP were funded by the Portuguese FCT through the LASIGE Strategic Project (UID/CEC/00408/2013). CP was also funded by FCT (PTDC/EEI-ESS/4633/2014). The research of IFC and BS was partially supported by a grant from the Bloomberg Philanthropies and by NSF awards CNS-1646395, III-1618126, CCF-1331800, III-1213013, and IIS-1143926.

References

1. G. Antunes, M. Bakhshandeh, J. Borbinha, J. Cardoso, S. Dadashnia, C. Francescomarino, M. Dragoni, P. Fettke, A. Gal, C. Ghidini, et al. The process model matching contest 2015. In *6th EMISA Workshop*, pages 127–155, 2015.
2. J. Cardoso, M. Bakhshandeh, D. Faria, C. Pesquita, and J. Borbinha. Ontology-Based Approach for Heterogeneity Analysis of EA Models. In *Workshop on Business Process Management and Ontologies*, 2016.
3. I. F. Cruz, F. Palandri Antonelli, and C. Stroe. AgreementMaker: Efficient Matching for Large Real-World Schemas and Ontologies. *PVLDB*, 2(2):1586–1589, 2009.
4. I. F. Cruz, C. Stroe, F. Caimi, A. Fabiani, C. Pesquita, F. M. Couto, and M. Palmonari. Using AgreementMaker to Align Ontologies for OAEI 2011. In *ISWC International Workshop on Ontology Matching (OM)*, volume 814 of *CEUR Workshop Proceedings*, pages 114–121, 2011.

5. D. Faria, C. Martins, A. Nanavaty, D. Oliveira, B. S. Balasubramani, A. Taheri, C. Pesquita, F. M. Couto, and I. F. Cruz. AML results for OAEI 2015. In *Ontology Matching Workshop*. CEUR, 2015.
6. D. Faria, C. Pesquita, E. Santos, M. Palmonari, I. F. Cruz, and F. M. Couto. The AgreementMakerLight Ontology Matching System. In *OTM Conferences - ODBASE*, pages 527–541, 2013.
7. C. J. Mungall, C. Torniai, G. V. Gkoutos, S. Lewis, and M. A. Haendel. Uberon, an Integrative Multi-species Anatomy Ontology. *Genome Biology*, 13(1):R5, 2012.
8. D. Oliveira and C. Pesquita. Compound matching of biomedical ontologies. In *International Conference on Biomedical Ontology*, volume 1515. CEUR, 2015.
9. C. Pesquita, D. Faria, C. Stroe, E. Santos, I. F. Cruz, and F. M. Couto. What's in a "nym"? Synonyms in Biomedical Ontology Matching. In *International Semantic Web Conference (ISWC)*, pages 526–541, 2013.
10. P. N. Robinson, S. Köhler, S. Bauer, D. Seelow, D. Horn, and S. Mundlos. The human phenotype ontology: a tool for annotating and analyzing human hereditary disease. *The American Journal of Human Genetics*, 83(5):610–615, 2008.
11. C. Rosse, J. L. Mejino Jr, et al. A reference ontology for biomedical informatics: the foundational model of anatomy. *Journal of biomedical informatics*, 36(6):478–500, 2003.
12. B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall, et al. The obo foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, 25(11):1251–1255, 2007.
13. C. L. Smith and J. T. Eppig. The mammalian phenotype ontology as a unifying standard for experimental and high-throughput phenotyping data. *Mammalian genome*, 23(9-10):653–668, 2012.